JOURNAL OF BACTERIOLOGY, June 1996, p. 3077–3084 Vol. 178, No. 11
0021-9193/96/$04.00+0

A New Type of Cohesin Domain That Specifically Binds the Dockerin Domain of the *Clostridium thermocellum* Cellulosome-Integrating Protein CipA

EMMANUELLE LEIBOVITZ AND PIERRE BÉGUIN*

Unité de Physiologie Cellulaire and URA 1300 CNRS, Département des Biotechnologies, Institut Pasteur, 75724 Paris Cedex 15, France

Received 26 October 1995/Accepted 21 March 1996

The cellulosome-integrating protein CipA, which serves as a scaffolding protein for the cellulolytic complex produced by *Clostridium thermocellum*, comprises a COOH-terminal duplicated segment termed the dockerin domain. This paper reports the cloning and sequencing of a gene, termed *sdbA* (for scaffoldin dockerin binding), encoding a protein which specifically binds the dockerin domain of CipA. The sequenced fragment comprises an open reading frame of 1,893 nucleotides encoding a 631-amino-acid polypeptide, termed SdbA, with a calculated molecular mass of 68,577 kDa. SdbA comprises an NH_2-terminal leader peptide followed by three distinct regions. The NH_2-terminal region is similar to the NH_2-terminal repeats of *C. thermocellum* OlpB and ORF2p. The central region is rich in lysine and harbors a motif present in *Streptococcus* M proteins. The COOH-terminal region consists of a triplicated sequence present in several bacterial cell surface proteins. The NH_2-terminal region of SdbA and a fusion protein carrying the first NH_2-terminal repeat of OlpB were shown to bind the dockerin domain of CipA. Thus, a new type of cohesin domain, which is present in one, two, and four copies in SdbA, ORF2p, and OlpB, respectively, can be defined. Since OlpB and most likely SdbA and ORF2p are located in the cell envelope, the three proteins probably participate in anchoring CipA (and the cellulosome) to the cell surface.

Clostridium thermocellum, a gram-positive thermophilic and anaerobic bacterium (26), produces a high-molecular-weight cellulase complex termed cellulosome (19). This complex is originally bound to the cell surface and subsequently released into the medium. The cellulosome is composed of at least 15 different polypeptides, including numerous β-1,4-endoglucanases (15), at least one cellobiohydrolase (28), and several hemicellulases (β-1,4-xylanases and lichenases) (27). Catalytic components are bound to a noncatalytic scaffolding subunit, named CipA (cellulosome integrating protein) (11, 41, 46). CipA and similar components identified in cellulase complexes of other cellulolytic clostridia have been termed scaffoldins (2).

The mode of attachment of the catalytic subunits to CipA has been elucidated. Each catalytic subunit contains a conserved duplicated segment of 23 residues (3), which has been termed the dockerin domain (2). The dockerin domains interact with a set of complementary binding domains of about 145 to 150 residues (8, 41), called cohesin domains (2), which are reiterated ninefold within the sequence of CipA (11). Besides the reiterated cohesin domains, the CipA polypeptide includes a cellulose-binding domain and a C-terminal duplicated segment of 23 residues (11). This segment resembles the dockerin domains of the catalytic subunits, although its amino acid sequence is more divergent from the consensus. The binding properties of the duplicated segment of CipA have been studied by using the chimeric protein CelC-DSCipA, in which the duplicated segment is fused to the COOH terminus of CelC (36). CelC-DSCipA binds neither to the cohesin domains of CipA nor to the cohesin domain of OlpA (previously termed ORF3p), a cell surface protein whose gene is located downstream from the *cipA* gene (9). However, CelC-DSCipA binds to three extracellular polypeptides of *C. thermocellum* with molecular masses of 170, 116, and 60 kDa, termed p170, p116, and p60, respectively (36). Thus, the duplicated segment of CipA behaves like a dockerin domain with a binding specificity differing from that of the dockerin domains borne by the catalytic subunits. The structures and functions of the polypeptides that bind the dockerin domain of CipA are unknown.

In this study, we report the molecular cloning and sequence analysis of a gene whose product specifically binds the dockerin domain borne by CelC-DSCipA. Segments of the gene were subcloned and expressed separately in order to identify the region of the polypeptide responsible for binding the dockerin domain of CipA. The identified region was similar to reiterated segments present in the previously described OlpB protein (9, 20). One of the OlpB segments was shown to bind CelC-DSCipA.

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture conditions. Bacterial strains and plasmids used in this study are described in Table 1. *Escherichia coli* TG1 was used for cloning and sequencing. Proteins were produced in *E. coli* M15(pREP4).

C. thermocellum was grown anaerobically at 60°C in CM3-3 medium supplemented with 5 g of cellobiose per liter (39).

E. coli was grown at 37°C in Luria-Bertani medium (23). The following antibiotics were added, depending on the plasmids present in the host: ticarcillin (100 μg/ml), chloramphenicol (30 μg/ml), and kanamycin (25 μg/ml).

DNA manipulations. *C. thermocellum* genomic DNA was purified by the method of Marmur as modified by Quiviger et al. (32). Other DNA manipulations were performed as described by Ausubel et al. (1). Restriction enzymes were used as recommended by the suppliers.

Oligonucleotides primers were synthesized by Eurogentec SA (Seraing, Belgium) or Genset SA (Paris, France). PCR amplification was performed as described by Saiki et al. (34), using 100 pmol of each oligonucleotide primer in 100 μl of reaction mix. $MgCl_2$ was added to a final concentration of 2 mM. Thirty-five amplification cycles were performed with the following parameters: annealing, 1 min at 65°C; elongation, 1 min at 72°C; and denaturation, 1 min at 94°C. The sequences of cloned PCR fragments were verified throughout.

Construction of the *C. thermocellum* genomic library. *C. thermocellum* DNA

* Corresponding author. Mailing address: Unité de Physiologie Cellulaire, Département des Biotechnologies, Institut Pasteur, 28, rue du Dr. Roux, 75724 Paris Cedex 15, France. Fax: 33 1 45 68 87 90. Electronic mail address: beguin@pasteur.fr.

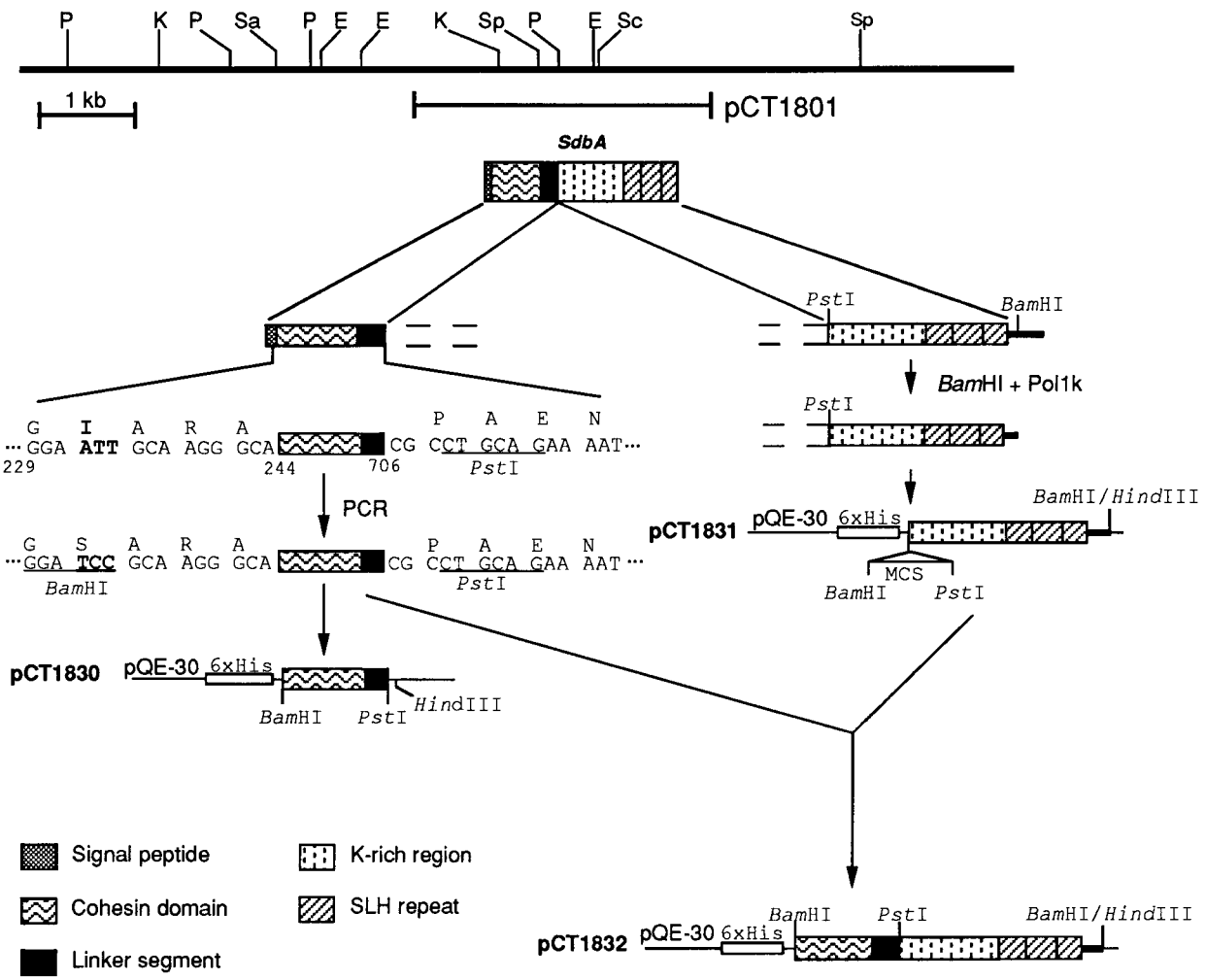


FIG. 1. Restriction map of the region including the *sdbA* gene and construction of pCT1830, pCT1831, and pCT1832, encoding SdbA-N, SdbA-C, and SdbA, respectively. E, *Eco*RI; K, *Kpn*I; P, *Pst*I; Sa, *Sal*I; Sc, *Sac*I; Sp, *Sph*I, MCS, multiple cloning site. The position of the insert carried by pCT1801 is indicated. The positions of the segments encoding the various regions identified within SdbA are shown by boxes of different patterns. Numbers refer to the nucleotide sequence (Fig. 3). Nucleotides that were changed in the PCR-amplified sequence are shown in boldface type. The DNA of the pQE-30 vector is indicated by a thin line. The pQE-30 sequence encoding six histidines is represented by a box not drawn to scale. The transcription of *sdbA* is from left to right.

was partially digested by *Sau*3AI, and fragments were separated on a sucrose gradient. Fragments larger than 12 kb were inserted in plasmid pUC18 cleaved by *Bam*HI and treated with bacterial alkaline phosphatase (Ready-to-Go; Pharmacia). *E. coli* TG1 cells were transformed by electroporation and plated in the presence of 0.8 mg of 5-bromo-4-chloro-3-indolyl-β-D-galactoside and 0.2 mg of isopropyl-β-D-thiogalactoside (IPTG) per plate.

Colony screening and Western blot analysis. CelC-DSCipA (20 μg) was labeled with 100 to 200 μCi of $Na^{125}I$, using H_2O_2 and lactoperoxidase (24, 41). Colonies of recombinant clones were overlaid overnight at 37°C with nitrocellulose filters (Hybond-C; Amersham). Filters were washed four times for 30 min each in phosphate-buffered saline (PBS; 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl [pH 7.3]) containing 5% nonfat dry milk and incubated overnight at room temperature in the same buffer containing 4×10^4 cpm of ^{125}I-labeled CelC-DSCipA per ml. Excess radioactivity was eliminated by six washes in the same buffer and then three washes in PBS. Filters were blotted dry, covered with plastic wrap, and autoradiographed at −70°C with an intensifying screen. Among positive clones, pCT1801 was studied in detail.

For Western blots, proteins were analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (16) and transferred by electrophoresis onto a nitrocellulose membrane (Hybond-C; Amersham) (42). Detection of proteins interacting with ^{125}I-labeled CelC-DSCipA or ^{125}I-labeled SdbA was performed as described above, using 2×10^5 cpm of the ^{125}I-labeled probe per ml.

DNA sequencing and sequence analysis. Appropriate restriction fragments of pCT1801 were subcloned in plasmid pBC SK^-, and nested deletions were generated by using exonuclease III and S1 nuclease (Erase-a-Base kit; Promega) as recommended by the supplier. Single-stranded templates were sequenced by the dideoxy-chain termination method (38), using Sequenase and Taquence kits (USB-Amersham). The sequence was determined at least once on each strand. Computer analysis of sequence data was performed with the sequence analysis software package (version 7) of the Genetics Computer Group (University of Wisconsin) (6).

Construction of expression clones and protein purifications. Clones overproducing intact or deleted forms of SdbA (scaffoldin dockerin binding) were constructed by using the pQE-30 vector. The sequence encoding the desired polypeptide was fused to a segment encoding six His residues to facilitate purification (14). To clone the fragment encoding the NH_2-terminal domain of SdbA, a 670-bp fragment flanked by *Bam*HI and *Pst*I sites was synthesized by PCR (Fig. 1). The forward primer was 5′-CTG CCG GCG GGA TCC GCA AGG GCA GAT-3′, and the reverse primer was 5′-ACT TTT GCA GAA TTT TCT GCA GGC G-3′. The fragment was inserted between the *Bam*HI and *Pst*I sites of pQE-30, yielding pCT1830. The polypeptide encoded by pCT1830 was called SdbA-N.

To clone the region encoding the COOH-terminal domains of SdbA, plasmid pCT1801 was digested with *Bam*HI. The ends were filled and converted into blunt ends with the Klenow fragment of DNA polymerase. After recutting with *Pst*I, the 1.4-kb fragment encoding the central and COOH-terminal regions of SdbA was purified and inserted into the pQE-30 vector which had been digested with *Hin*dIII, treated with the Klenow fragment of DNA polymerase, and digested again with *Pst*I. The resulting plasmid was termed pCT1831, and the encoded polypeptide was called SdbA-C.

Plasmid pCT1832, expressing the sequence of SdbA (without signal peptide),

TABLE 1. Bacterial strains and plasmids used

Strain or plasmid	Relevant features	Reference or source
Strains		
E. coli		
TG1	[Δ(*lac-pro*) *thi supE hsdD5*/F′ *tra-36proA*$^+$*B*$^+$*lacI*^{q} *lacZ*ΔM15]	13
M15(pREP4)		7, 44; Qiaexpress kit (Qiagen)
C. thermocellum		
NCIB 10682		
YS		19; R. Lamed
Plasmids		
pUC18	Amp^r, cloning vector	47
pBC SK$^-$	Cam^r, cloning vector	Stratagene
pQE-30	Amp^r, expression vector	Qiaexpress kit (Qiagen)
pCT1801	pUC18 derivative containing an *Sau*3A fragment encoding SdbA	This study
pCT1830	pQE-30 derivative encoding the cohesin domain of SdbA fused to 6 His residues	This study
pCT1831	pQE-30 derivative encoding the central and COOH-terminal regions of SdbA fused to 6 His residues	This study
pCT1832	pQE-30 derivative encoding SdbA fused to 6 His residues	This study

was constructed by inserting the 670-bp *Bam*HI-*Pst*I fragment (see above) into plasmid pCT1831 digested by *Bam*HI and *Pst*I.

Production and purification of proteins were performed with the Qiaexpress system (Qiagen Inc). One-liter cultures were grown at 37°C to an optical density at 600 nm of 0.7. IPTG was then added to a final concentration of 0.3 mM, and the cultures were further incubated at 37°C for 5 h. Cells were resuspended in 80 ml of 50 mM Tris-HCl (pH 7.5) (buffer A) and disrupted in an Aminco French pressure cell at 14,000 lb/in^2 (100 MPa). The extract was centrifuged at 9,000 × *g* for 20 min to remove cell debris. The supernatant was loaded on an 8-ml column of Ni-nitrilotriacetic acid resin equilibrated with buffer A, washed with buffer A, and eluted with the same buffer containing 250 mM imidazole. The eluted fractions were dialyzed overnight at 4°C against 1 liter of buffer A. Purified proteins were stored at −80°C.

Amino-terminal amino acid sequence determination. Fifty picomoles of each polypeptide to be sequenced was subjected to SDS-PAGE and transferred overnight, at room temperature, at 850 mA onto a hydrophobic polyvinylidene difluoride membrane (ProBlott; Applied Biosystems) treated with 100% methanol, using a Trans-Blot Cell (Bio-Rad) containing 50 mM Tris base and 50 mM boric acid buffer. Bands were stained with 0.003% amido black and cut out, and the amino-terminal sequences of the polypeptides were determined by the Edman method, using a model 473A or a Procise HT sequencer (Applied Biosystems).

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper will appear in the GenBank nucleotide sequence database under accession number U49980.

RESULTS

Cloning of a gene encoding a polypeptide which specifically binds the dockerin domain of CipA. We screened 1,600 recombinant clones for the binding of ^{125}I-labeled CelC-DSCipA. Eight independent clones were labeled. Controls performed with ^{125}I-labeled CelC-DSCelD, which carries the dockerin domain of CelD (40), indicated that binding was specific for the dockerin domain of CipA (Fig. 2).

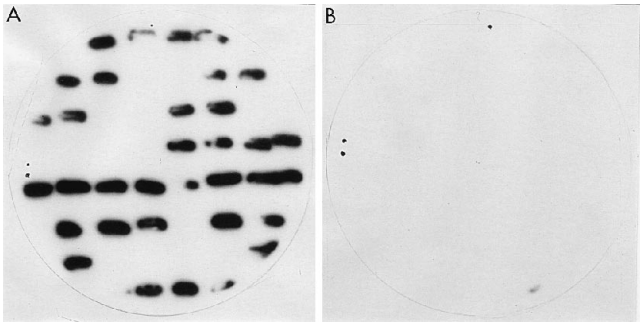
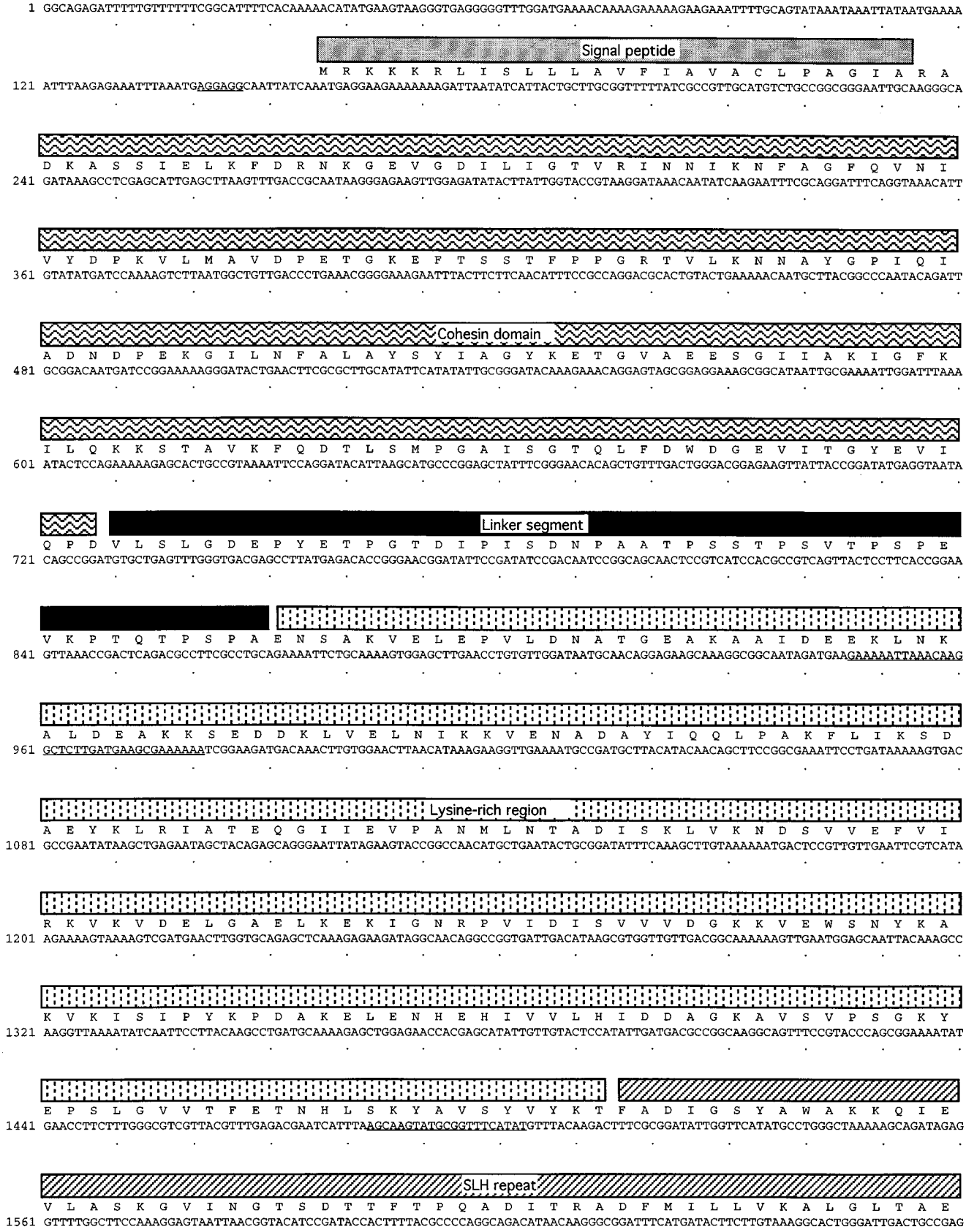


FIG. 2. Specific binding of ^{125}I-labeled CelC-DSCipA to clones producing SdbA. Positive and negative clones patched on duplicate plates were transferred onto nitrocellulose filters, and the filters were incubated with ^{125}I-labeled CelC-DSCipA (A) or ^{125}I-labeled CelC-DSCelD (B). Filters were washed and autoradiographed as described previously (8).

All cloned segments hybridized to the same region of the *C. thermocellum* genome (data not shown), whose map is shown in Fig. 1. Their restriction maps were consistent with the restriction fragments revealed by Southern blotting in *C. thermocellum* DNA (data not shown). The segments did not hybridize (data not shown) and did not have restriction fragments in common with the region extending between *cipA* and *olpA* (9). Within the region covered by the cloned fragments, a 1.6-kb segment between the *Pst*I site and the left boundary of the insert borne by pCT1801 (Fig. 1) was necessary and sufficient to encode a polypeptide able to bind the dockerin domain of CipA. The corresponding gene was termed *sdbA* (for scaffoldin dockerin binding).

Sequence analysis. The sequence of the *sdbA* gene is shown in Fig. 3. The coding sequence comprises 1,893 nucleotides. The ATG start codon is preceded by a putative ribosome-binding site. The encoded polypeptide, composed of 631 amino acids, has a calculated molecular mass of 68,577 Da. The domain structure of the protein is shown in Fig. 1 and 3. A putative signal peptide of 26 amino acid residues is located at the NH_2 terminus of the polypeptide (45). Alignments with other proteins indicated the presence of three distinct regions in SdbA. The N-terminal region, composed of 156 amino acid residues, is similar to the N-terminal repeats of *C. thermocellum* OlpB (previously termed ORF1p) and ORF2p, two polypeptides whose genes are located immediately downstream from *cipA* (9) (Fig. 4). A Pro/Thr/Ser-rich spacer of 56 residues separates this region from the rest of the protein. The central region is composed of 215 amino acids, with many Lys residues. This region comprises a short amino acid sequence similar to a segment present in *Streptococcus pyogenes* M proteins (Fig. 5). The COOH-terminal region is composed of three repeats that are highly similar to the segments termed SLH (S-layer homologous) present in several proteins located on the cell surface of various bacteria (9, 22) (Fig. 6).

Identification of the domain responsible for binding the dockerin domain of CipA. To identify the domain responsible for binding the dockerin domain of CipA, the binding properties of polypeptides derived from SdbA were compared. The *sdbA* gene and appropriate subfragments were fused to the His_6-encoding expression vector pQE-30, and the corresponding polypeptides were purified by Ni^{2+} affinity chromatogra-


Signal peptide
Cohesin domain
Linker segment
Lysine-rich region
SLH repeat

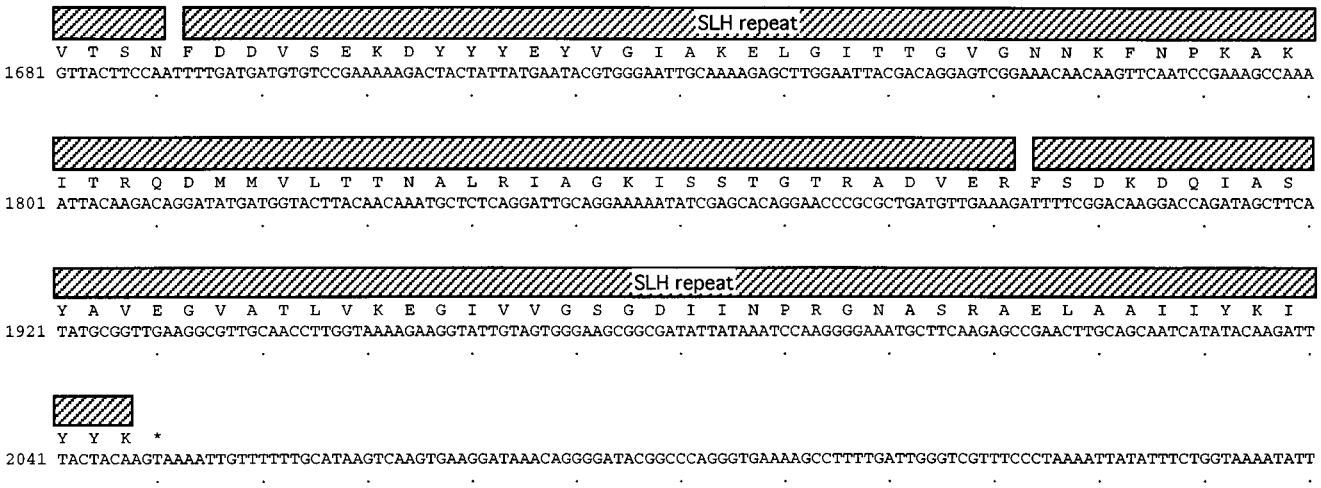


FIG. 3. Nucleotide sequence of the region encoding the *sdbA* gene. The putative ribosome-binding site, the motif found in streptococcal M proteins (Fig. 5), and the NH_2 terminus of the 24-kDa species present in lanes 2 and 3 of Fig. 7A are underlined. The various regions identified within SdbA are indicated by boxes with the same patterns as in Fig. 1.

phy. The apparent molecular masses of intact SdbA and of the fragment containing the central and C-terminal regions were 60 and 36 kDa, respectively, in agreement with the masses predicted from the nucleotide sequence (Fig. 7A). The apparent molecular mass of the NH_2-terminal region was 35 kDa, higher than the molecular mass calculated from the nucleotide sequence (22,715 Da). However, the Pro-rich linker represents a large fraction of the polypeptide, which may account for a slow migration in SDS-PAGE (10). The preparations of intact SdbA and of the COOH-terminal polypeptide both contained a second polypeptide of 24 kDa. In both cases, the NH_2-terminal sequence of the 24-kDa species was SKYAVSY, a segment located at the end of the lysine-rich region. This finding indicates that the 24-kDa polypeptide was derived from the COOH-terminal region containing the SLH repeats of SdbA. Since SLH repeats contain no histidine cluster, the COOH-terminal fragment was probably bound to the intact polypeptides. Indeed, polypeptides containing SLH repeats have been reported to self-associate (20).

Western blotting with ^{125}I-labeled CelC-DSCipA as a probe confirmed that the product of the *sdbA* gene bound the dockerin domain of CipA (Fig. 7B). Binding to the NH_2-terminal fragment was weak but detectable. Binding to the C-terminal fragment could not be detected even after a sevenfold-longer exposure time (data not shown). No binding of ^{125}I-labeled CelC-DSCipA to SdbA was observed when Ca^{2+} (present in nonfat dry milk added as a blocking agent) was chelated in the presence of 25 mM EDTA (data not shown), which suggests that the interaction was dependent on the presence of Ca^{2+} (or some other cation chelated by EDTA). A similar observation was made by Yaron et al. (48) for the binding of the second and third cohesin domains of CipA to the catalytic components of the cellulosome.

Binding of ^{125}I-labeled SdbA to CipA. Western blotting with ^{125}I-labeled SdbA as a probe was performed with purified recombinant CipA and with various CipA-containing fractions obtained from cultures of *C. thermocellum* (Fig. 8). In each case, the migration of the topmost band corresponded to an apparent molecular mass of 210 to 250 kDa, in agreement with previous results reported for the scaffoldins of *C. thermocellum* NCIB 10682 and YS. The scaffoldin of strain YS, termed CipB (31), migrates somewhat faster than CipA from strain ATCC 27405 (=NCIB 10682) (17). In addition, each sample displayed a set of regularly spaced bands migrating faster than CipA. The bands probably derived from proteolytic cleavages occurring at regularly spaced intervals along the scaffoldin sequence (most likely in the linker regions between the cohesin domains). CipA is highly sensitive to proteolysis, and ladders were observed previously for CipA prepared from the cellulosome by preparative electrophoresis (37). Most of the labeled polypeptides detected in either strain YS or strain NCIB 10682 were identical in size, but CipA seemed to contain one more module than CipB. Since both CipA and CipB contain a cellulose-binding domain (11, 29), it is tempting to speculate that CipB has one cohesin domain less than CipA.

Binding of CelC-DSCipA to the NH_2-terminal repeat of *C. thermocellum* OlpB. Since the NH_2-terminal region of SdbA is similar to the NH_2-terminal repeats of OlpB, we checked whether CelC-DSCipA would bind to MalE-ORF1p-N, a chimeric protein comprising the first NH_2-terminal repeat of OlpB fused to the maltose-binding protein MalE (20). MalE-ORF1p-N appeared to be labeled (Fig. 7B, lane 5). No binding was observed with MalE-ORF1p-C, which consists of the C-

```
 27 RADKASSIELKFDRNKGEVGDILIGTVRINNIKNFAGFQVNIVYDPKVLMAVDPET  SdbA
 28 .AEATPSIEMVLDKTEVHVGDVITATIKVNNIRKLAGYQLNIKFDPEVLQPVDPAT  OlpB
207 .........LELDKTKVKVGDIITATIKIENMKNFAGYQLNIKYDPTMLEAIELET  OlpB
409 .........MELDKTKVKVGDIITATIKIENMKNFAGYQLNIKYDPTMLEAIELET  OlpB
607 .........MELDKTKVKEGDVIIATIRVNNIKNLAGYQIGIKYDPKVLEAFNIET  OlpB
 38 .........MELDKTKANIGDIIIATIRIDNINNFSGYQLNIKYDPSYLQAVNPLT  ORF2p
209 ........ALELDKTKVKVGDVIVATVKAKNMTSMAGIQVNIKYDPEVLQAIDPAT  ORF2p

 83 GKEFTSSTFPP..GRTVLKNNAYGPIQIADNDPEKGILNFALAYSYIAGYKETGVA  SdbA
 84 GEEFTDKSMPV..NRVLLTNSKYGPTPVAGNDIKSGIINFATGYNNLTAYKSSGID  OlpB
254 GSAIAKRTWPVTGGTV.LQSDNYGKTTAVANDVGAGIINFAEAYSNLTKYRETGVA  OlpB
456 GSAIAKRTWPVTGGTV.LQSDNYGKTTAVANDVGAGIINFAEAYSNLTKYRETGVA  OlpB
654 GDPIDEGTWPAVGGTI.LKNRDYLPTGVAINNVSKGILNFAAYYVYFDDYREEGKS  OlpB
 85 GEPIKKRTMPAVNGTVLLKGDQYSITEVVENNVDEGILNFGKGYANLTEYRKSGKP  ORF2p
257 GKPFTKETLLV..DPELLSNREYNPLLTAVNDINSGIINYASCYVYWDSYRESGVS  ORF2p

137 EESGIIAKIGFKILQKKSTAVKFQDTLSMPGAISGTQLFDWDGEVITGYEVIQP  SdbA
138 EHTGIIGEIGFKVLKKQNTSIRFEDTLSMPGAISGTSLFDWDAETITGYEVIQP  OlpB
309 EETGIIGKIGFRVLKAGSTAIRFEDTTAMPGAIEGTYMFDWYGENIKGYSVVQP  OlpB
511 EETGIIGKIGFRVLKAGSTAIRFEDTTAMPGAIEGTYMFDWYGENIKGYSVVQP  OlpB
709 EDTGIIGNIGFRVLKAEDTTIRFEELESMPGSIDGTYMFDWYLNRISGYVVIQP  OlpB
141 ETTGIIGKIGFKALKLGKTEIKFENTPVMPGAKEGTLLFDWDAETITEYNVIQP  ORF2p
311 ESTGIIGKVGFKVLKAANTTVKLEETRFTPNSIDGTLVIDWYGQQIVGYKVIQP  ORF2p
```

FIG. 4. Alignment of the cohesin domain of SdbA and of the NH_2-terminal repeats of OlpB and ORF2p (9). Residues that are identical or similar in the majority of the displayed sequences are shown against a shaded background. Numbering of residues starts with putative initiation codons. Similar amino acids are F, I, V, L, and M; R and K; S and T; D and E; N and Q; and F, Y, and W.

```
 98 EKAKQALEDQRK M1
264 EKLNKALDEAKK SdbA
278 EKLNKELEEGKK M9
289 EKLNKELEESKK PAM
450 EKLNKDLEESKK M12
```

FIG. 5. Similarity between residues 264 to 275 of SdbA and a motif present in *S. pyogenes* M proteins M1 (GenBank accession number X72752), M9 (30), PAM (4), and M12 (33). For each protein, numbering starts with the putative initiation codon. Residues that are identical or similar in the majority of the displayed sequences are shown against a shaded background. Similarity criteria are the same as for Fig. 4.

terminal SLH segments of OlpB fused to MalE (lane 4). Neither SdbA, MalE-ORF1p-N, nor MalE-ORF1p-C was labeled after incubation with ^{125}I-labeled CelC-DSCelD (data not shown).

DISCUSSION

Proteins of *C. thermocellum* carrying dockerin domains can be labeled with ^{125}I and used as probes to detect proteins harboring complementary cohesin domains (36, 40, 41). Thus, clones expressing cohesin-containing polypeptides can be isolated, and the cohesin domains can be identified (8). In this study, the same strategy was applied to clone the *sdbA* gene and to identify the cohesin domain responsible for binding the dockerin domain of CipA. Only one gene was obtained. It may be that other genes encoding proteins with similar properties escaped detection because of lack of expression in *E. coli.*

Of the three polypeptides, p170, p116, and p60, which were previously shown to bind the dockerin domain of CipA (36), p170 and p116 are probably too large to be encoded by *sdbA*, even allowing for posttranslational modifications such as glycosylation. The p60 polypeptide appears as a more likely candidate; however, positive identification of the SdbA polypeptide as p60 requires further confirmation, for example, by demonstrating immunological cross-reactivity.

Figure 7 indicates that the cohesin domain lies in the NH_2-proximal region of SdbA. The signal detected with the NH_2-terminal fragment was much weaker than that detected with the whole protein. However, truncation of SdbA may affect the affinity or stability of the residual NH_2-terminal polypeptide. Alternatively, binding to nitrocellulose may alter the conformation of the isolated cohesin domain, whereas attachment of the intact protein to the membrane may be mediated by regions of the polypeptide not required for binding the labeled probe.

In contrast to the dockerin domain of CipA, which is clearly related to the dockerin domains present in the catalytic subunits (8, 11), the cohesin domain of SdbA shows no obvious similarity with the cohesin domains of CipA and OlpA. However, it is similar to the repeats located at the NH_2-terminal ends of OlpB and ORF2p (9). Indeed, ^{125}I-labeled CelC-DSCipA bound specifically to the first NH_2-terminal repeat of OlpB. Thus, the NH_2-terminal domains of SdbA, OlpB, and most likely ORF2p represent a new type of cohesin domain. We propose to classify these cohesin domains as type II and to classify as type I the cohesin domains found in CipA and at the NH_2 terminus of OlpA. The same distinction between type I and type II applies to the cognate dockerin domains present in catalytic cellulosome components and in CipA, respectively.

The three proteins, SdbA, OlpB, and ORF2p, that are known to contain cohesin domains of type II also carry SLH repeats. In all cases studied so far, SLH repeats are found in proteins that are associated with the cell surface of bacteria, and biochemical evidence indicates that they bind to components of the cell envelope (20). Thus, SdbA may be located on the cell surface, like OlpA (35) and OlpB (20). The similarity between the central region of SdbA and a region present in streptococcal M proteins reinforces the hypothesis. It has been surmised that in M proteins, this region may interact with cell wall carbohydrates (43). Taken together, these considerations strongly suggest that SdbA, OlpB, and possibly ORF2p are components of the cell envelope that are involved in binding cellulosomes to the cell surface.

The structural organization of the cellulosome and the attachment of the cellulosome to the cell surface depend on the same type of protein-protein interactions. In both cases, the interaction involves the binding of a duplicated 24-amino-acid segment, termed the dockerin domain, to a segment of 140 to 160 amino acids termed the cohesin domain. Cohesin domains of type I, such as those found in CipA, are known to bind to a variety of dockerin domains of type I borne by catalytic components of the cellulosome. By contrast, the dockerin domain of CipA is the only dockerin domain of type II known to bind to cohesin domains of type II. Indeed, *cipA* and *olpB* are cotranscribed (9), emphasizing the likelihood that their products have related functions.

The general structures of OlpB, ORF2p, and SdbA fit well with a role in displaying proteins toward the outside of the cell. The cohesin domains are located at the NH_2 terminus of the polypeptides, on the opposite end from the COOH-terminal SLH segments, which are bound to the cell surface. Accordingly, on whole cells, the cohesin domains of OlpB are more easily labeled with antibodies than the SLH segments (20).

Why three different proteins (and possibly more, if p116 and p170 belong to the same family) should be needed to display cellulosomes on the cell surface is open to speculation. SdbA,

```
1453 .....AYLRGY...PDGSFRPERNITRAEAAVIFAKLLGADESYGAQSASP........  OlpB
1496 YSDLAD.THWAAWAIKFATSQGLFKGY...PDGTFKPDQNITRAEFATVVLHFLTKVKGQEIMSKLATIDISNPK  OlpB
1567 FDDCVG..HWAQEFIEKLTSLGYISGY...PDGTFKPQNYITRSESVALINRALERGPLNGAPKL.........  OlpB
1627 FPDVNE.SYWAFGDIMDGALD                                                      OlpB
 241 .....PFLKGY...PGGLFKPENNITRAEAAVIFAKLLGADENSAGKNSSIT.......  OlpA
 285 FKDLKD.SHWAAWAIKYVTEQNLFGGY...PDGTFMPDKSITRAEFATVTYKFLEKLGKIEQGTDVKTQ......  OlpA
 350 LKDIEG..HWAQKYIETLVAKGYIKGY...PDETFRPQASITKRAESVALINRSLERGPLNGAVLE.........  OlpA
 410 FTDVPV.NYWAYKDIAEGVIY                                                     OlpA
 454 FADIGS.YAWAKKQIEVLASKGVING...TSDTTFTPQADITRADFMILLVKALGLTAEVTSN...........  SdbA
 513 FDDVSE.KDYYYEYVGIAKELGITTG...VGNNKFNPKAKITRQDMMVLTTNALRIAGKISSTGTRADVER....  SdbA
 580 FSDKDQIASYAVEGVATLVKEGIVVG....SGDIINPRGNASRAELAAIIYKIYYK                    SdbA
1682 FNDIKD..NWAKDVIEVLASRHIVEG...MTDTQYEPNKTVTRAEFTAMILRLLNIKDETYSGE..........  Pul
1741 FSDVKS.GDWYANAIEAAYKAGIIEG...DGKNA.RPNDSITREEMTAIAMRAYEMLTQYKEENIGATT......  Pul
1805 FSDDKSISDWARNVVANAAKLGIVNG...EPNNVFAPKGNATRAEAAAIIYGLLEKSGNI..............  Pul
  36 LNDFNKISGYAKEAVQSLVDAGVIQG...DANGNFNPLKTISRAEAATIFTNALELEAEGDVN...........  Bsph
  93 FKDVKA.DAWYYDAIAATVENGIFEG...VSATEFAPNKQLTRSEAAKILVDAFELEGEGDLSE..........  Bsph
 153 FADASTVKPWAKSYLEIAVANGVIKGSEANGKTNLNPNAPITRQDFAVVFSRTIENVDATPKVDKIE.......  Bsph
```

FIG. 6. Alignment of the COOH-terminal repeats of SdbA with similar sequences of other cell surface proteins. OlpA, outer layer protein A of *C. thermocellum* (9, 35); OlpB, outer layer protein B of *C. thermocellum* (9, 20); Pul, pullulanase of *Thermoanaerobacterium thermosulfurigenes* EM1 (25); Bsph, S-layer protein of *Bacillus sphaericus* (5). For each protein, numbering starts at the putative initiation codon. Residues that are similar or identical in the majority of the duplicated sequences are shown against a shaded background. Similarity criteria are the same as for Fig. 4.

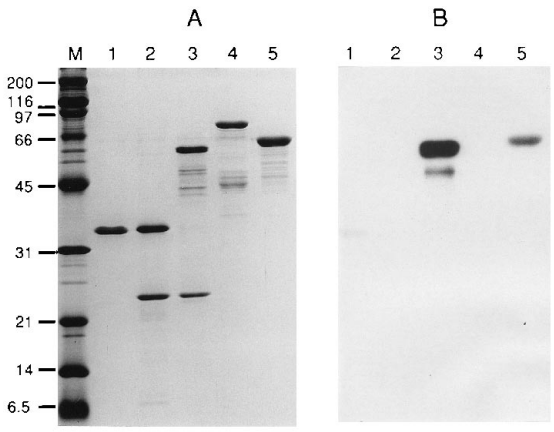


FIG. 7. Interaction of ^{125}I-labeled CelC-DSCipA with purified proteins. Proteins were loaded on an SDS–12.5% polyacrylamide gel and were either stained with Coomassie blue (A) or blotted onto a nitrocellulose filter and probed with ^{125}I-labeled CelC-DSCipA (B). Lanes: M, molecular mass markers (masses are indicated in kilodaltons); 1, SdbA-N; 2, SdbA-C; 3, SdbA; 4, MalE-ORF1p-C (20); 5, MalE-ORF1p-N (20).

ORF2p, and OlpB carry one, two, and four cohesin domains, respectively. The nature and the length of the segment connecting the SLH domains to the cohesin domains are also variable. OlpB features a connecting segment remarkably long and reiterated compared with those of the two other proteins. Possibly these differences are correlated to different roles in the topological organization of the complexes on the cell surface, in particular the formation of polycellulosomes.

ACKNOWLEDGMENTS

We thank R. Lamed for providing *C. thermocellum* YS, J. d'Alayer for performing amino acid sequences analyses, S. Salamitou and M. Lemaire for providing samples of CelC-DSCipA and MalE-ORF1p-N, respectively, and I. Miras for skillful technical assistance. We are grateful to J.-P. Aubert and M. Schwartz for their continuing interest and support.

E.L. was the recipient of a fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche. This work was supported by contract AIR1-CT-0321 from the Commission of the European Communities.

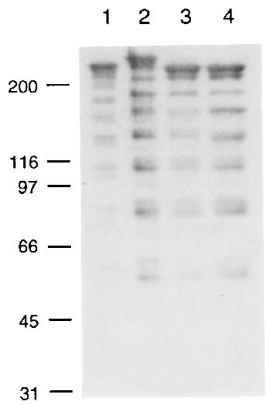


FIG. 8. Interaction of SdbA with various preparations containing CipA. Samples were loaded on a 5 to 15% polyacrylamide gel, transferred onto nitrocellulose, and probed with ^{125}I-labeled SdbA. Lanes: 1, purified CipA obtained from a recombinant *E. coli* clone (unpublished results); 2, cellulosome preparation purified by cellulose affinity chromatography (18) from the supernatant of a *C. thermocellum* NCIB 10682 culture; 3, supernatant from a *C. thermocellum* YS culture; 4, cell-bound proteins, prepared as described in reference 20, of *C. thermocellum* YS. The sizes (in kilodaltons) molecular mass markers are indicated at the left.

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1990. Current protocols in molecular biology. Greene Publishing and Wiley Interscience, New York.
2. **Bayer, E. A., E. Morag, and R. Lamed.** 1994. The cellulosome—a treasure-trove for biotechnology. Trends Biotechnol. **12:**379–386.
3. **Béguin, P., J. Millet, and J.-P. Aubert.** 1992. Cellulose degradation by *Clostridium thermocellum*: from manure to molecular biology. FEMS Microbiol. Lett. **100:**523–528.
4. **Berge, A., and U. Sjobring.** 1993. PAM, a novel plasminogen-binding protein from *Streptococcus pyogenes*. J. Biol. Chem. **268:**25417–25424.
5. **Bowditch, R. D., P. Baumann, and A. A. Yousten.** 1989. Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer protein from *Bacillus sphaericus* 2362 and a related cryptic gene. J. Bacteriol. **171:**4178–4188.
6. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
7. **Farabaugh, P. J.** 1978. Sequence of the *lacI* gene. Nature (London) **274:**765–769.
8. **Fujino, T., P. Béguin, and J.-P. Aubert.** 1992. Cloning of a *Clostridium thermocellum* DNA fragment encoding polypeptides that bind the catalytic components of the cellulosome. FEMS Microbiol. Lett. **94:**165–170.
9. **Fujino, T., P. Béguin, and J.-P. Aubert.** 1993. Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in the attachment of the cellulosome to the cell surface. J. Bacteriol. **175:**1891–1899.
10. **Furthmayr, H., and R. Timpl.** 1971. Characterization of collagen peptides by sodium dodecylsulfate-polyacrylamide electrophoresis. Anal. Biochem. **41:**510–516.
11. **Gerngross, U. T., M. P. M. Romaniec, N. S. Huskisson, and A. L. Demain.** 1993. Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal S_L-protein reveals an unusual degree of internal homology. Mol. Microbiol. **8:**325–334.
12. **Gibson, T. J.** 1984. Studies on the Epstein-Barr virus genome. Ph.D. thesis. University of Cambridge, Cambridge.
13. **Janknecht, R., G. de Martynoff, J. Lou, R. A. Hipskind, A. Nordheim, and H. G. Stunnenberg.** 1991. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proc. Natl. Acad. Sci. USA **88:**8972–8976.
14. **Kohring, S., J. Wiegel, and F. Mayer.** 1990. Subunit composition and glycosidic activities of the cellulase complex from *Clostridium thermocellum* JW20. Appl. Environ. Microbiol. **56:**3798–3804.
15. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
16. **Lamed, R., and E. A. Bayer.** 1988. The cellulosome concept: exocellular/extracellular enzyme reactor centers for efficient binding and cellulolysis. FEMS Symp. **43:**101–116.
17. **Lamed, R., E. Setter, and E. A. Bayer.** 1983. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. J. Bacteriol. **156:**828–836.
18. **Lamed, R., E. Setter, R. Kenig, and E. A. Bayer.** 1983. The cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. Biotechnol. Bioeng. Symp. **13:**163–181.
19. **Lemaire, M., H. Ohayon, P. Gounon, T. Fujino, and P. Béguin.** 1995. OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domain to components of the cell envelope. J. Bacteriol. **177:**2451–2459.
20. **Lupas, A., H. Engelhardt, J. Peters, U. Santarius, S. Volker, and W. Baumeister.** 1994. Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. J. Bacteriol. **176:**1224–1233.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Marchalonis, J. J.** 1969. An enzymic method for the trace iodination of immunoglobulins and other proteins. Biochem. J. **113:**299–305.
23. **Matuschek, M., G. Burchhardt, K. Sahm, and H. Bahl.** 1994. Pullulanase of *Thermoanaerobacter thermosulfurigenes* EM1 (*Clostridium thermosulfurogenes*): molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface. J. Bacteriol. **176:**3295–3302.
24. **McBee, R. H.** 1948. The culture and physiology of a thermophilic cellulose-fermenting bacterium. J. Bacteriol. **56:**653–663.
25. **Morag, E., E. A. Bayer, and R. Lamed.** 1990. Relationship of cellulosomal and non-cellulosomal xylanases of *Clostridium thermocellum* to cellulose-degrading enzymes. J. Bacteriol. **172:**6098–6105.
26. **Morag, E., I. Halevy, E. A. Bayer, and R. Lamed.** 1991. Isolation and properties of a major cellobiohydrolase from the cellulosome of *Clostridium*

thermocellum. J. Bacteriol. **173:**4155–4162.
27. **Morag, E., A. Lapidot, D. Govorko, R. Lamed, M. Wilchek, E. A. Bayer, and Y. Shoham.** 1995. Expression, purification, and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of *Clostridium thermocellum*. Appl. Environ. Microbiol. **61:**1980–1986.
28. **Podbielski, A., J. Hawlitzky, T. D. Pack, A. Flosdorff, and M. D. Boyle.** 1994. A group A streptococcal Enn protein potentially resulting from intergenomic recombination exhibits atypical immunoglobulin-binding characteristics. Mol. Microbiol. **12:**725–736.
29. **Poole, D. M., E. Morag, R. Lamed, E. A. Bayer, G. P. Hazlewood, and H. J. Gilbert.** 1992. Identification of the cellulose-binding domain of the cellulosome subunit S1 from *Clostridium thermocellum* YS. FEMS Microbiol. Lett. **99:**181–186.
30. **Quiviger, B., C. Franche, G. Lutfalla, R. D., R. Haselkorn, and C. Elmerich.** 1982. Cloning of a nitrogen fixation (*nif*) gene cluster of *Azospirillum brasilense*. Biochimie **64:**495–502.
31. **Robbins, J. C., J. G. Spanier, S. J. Jones, W. J. Simpson, and P. P. Cleary.** 1987. *Streptococcus pyogenes* type 12 M protein gene regulation by upstream sequences. J. Bacteriol. **169:**5633–5640.
32. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487–491.
33. **Salamitou, S., M. Lemaire, T. Fujino, H. Ohayon, P. Gounon, P. Béguin, and J.-P. Aubert.** 1994. Subcellular localization of *Clostridium thermocellum* ORF3p, a protein carrying a receptor for the docking sequence borne by the catalytic components of the cellulosome. J. Bacteriol. **176:**2828–2834.
34. **Salamitou, S., O. Raynaud, M. Lemaire, M. Coughlan, P. Béguin, and J.-P. Aubert.** 1994. Recognition specificity of the duplicated segments present in *Clostridium thermocellum* endoglucanase CelD and in the cellulosome-integrating protein CipA. J. Bacteriol. **176:**2822–2827.
35. **Salamitou, S., K. Tokatlidis, P. Béguin, and J. P. Aubert.** 1992. Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. FEBS Lett. **304:**89–92.
36. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
37. **Tailliez, P., H. Girard, J. Millet, and P. Béguin.** 1989. Enhanced cellulose fermentation by an asporogenous and ethanol-tolerant mutant of *Clostridium thermocellum*. Appl. Environ. Microbiol. **55:**207–211.
38. **Tokatlidis, K., P. Dhurjati, and P. Béguin.** 1993. Properties conferred on *Clostridium thermocellum* endoglucanase CelC by grafting the duplicated segment of endoglucanase CelD. Protein Eng. **6:**947–952.
39. **Tokatlidis, K., S. Salamitou, P. Béguin, P. Dhurjati, and J.-P. Aubert.** 1991. Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. FEBS Lett. **291:**185–188.
40. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76:**4350–4354.
41. **Vijaykumar, P., and V. A. Fischetti.** 1988. Isolation and characterization of the cell-associated region of group A streptococcal M6 proteins. J. Bacteriol. **170:**2618–2624.
42. **Villarejo, M. R., and I. Zabin.** 1974. β-galactosidase from termination and deletion mutant strains. J. Bacteriol. **120:**466–474.
43. **von Heijne, G.** 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. **133:**17–21.
44. **Wu, J. H. D., and A. L. Demain.** 1988. Proteins of the *Clostridium thermocellum* cellulase complex responsible for degradation of crystalline cellulose. FEMS Symp. **43:**117–131.
45. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
46. **Yaron, S., E. Morag, E. A. Bayer, R. Lamed, and Y. Shoham.** 1995. Expression, purification and subunit-binding properties of cohesins 2 and 3 of the *Clostridium thermocellum* cellulosome. FEBS Lett. **360:**121–124.